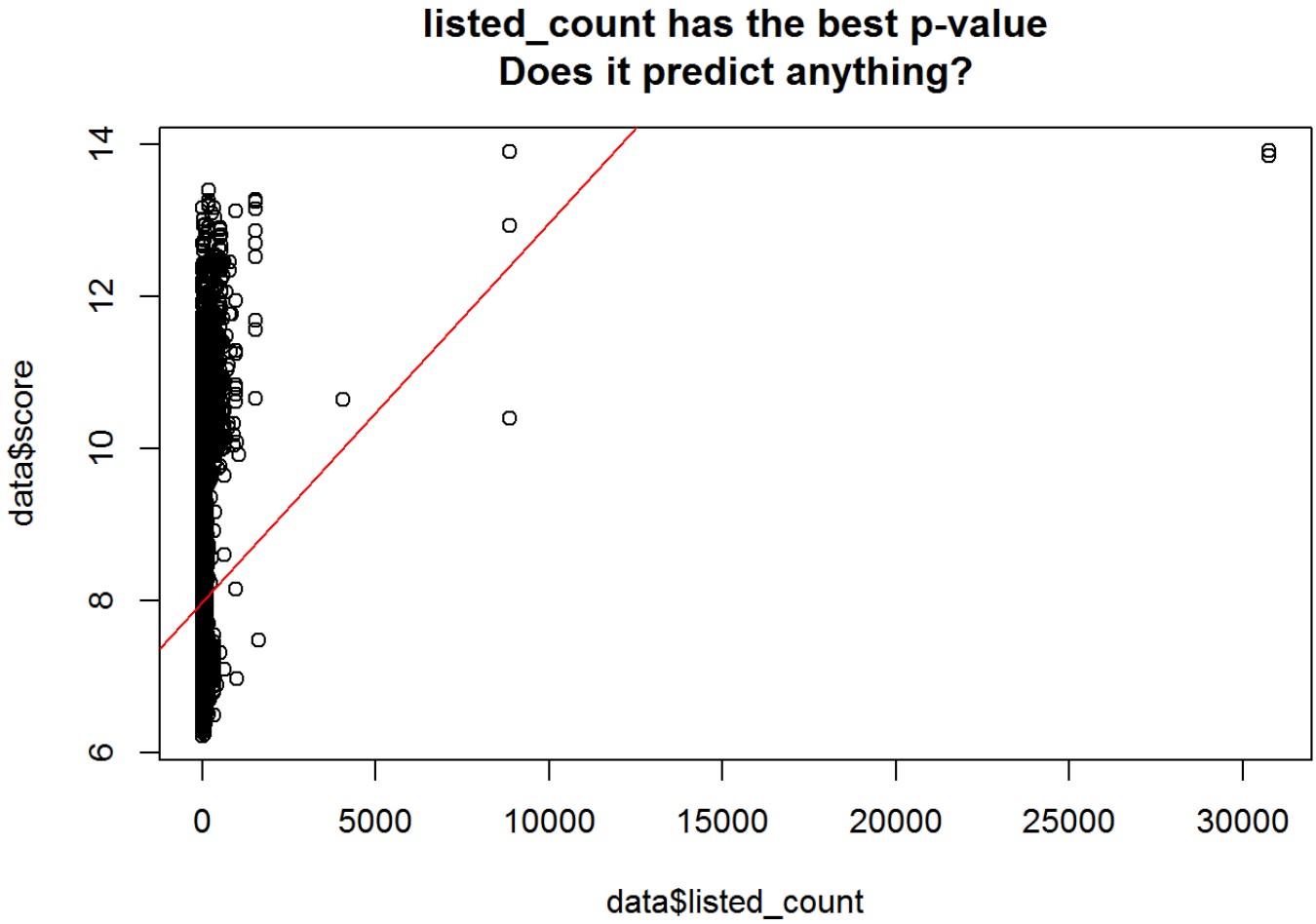
Healthcare Twitter Analytics

Do any of the numerics predict score?

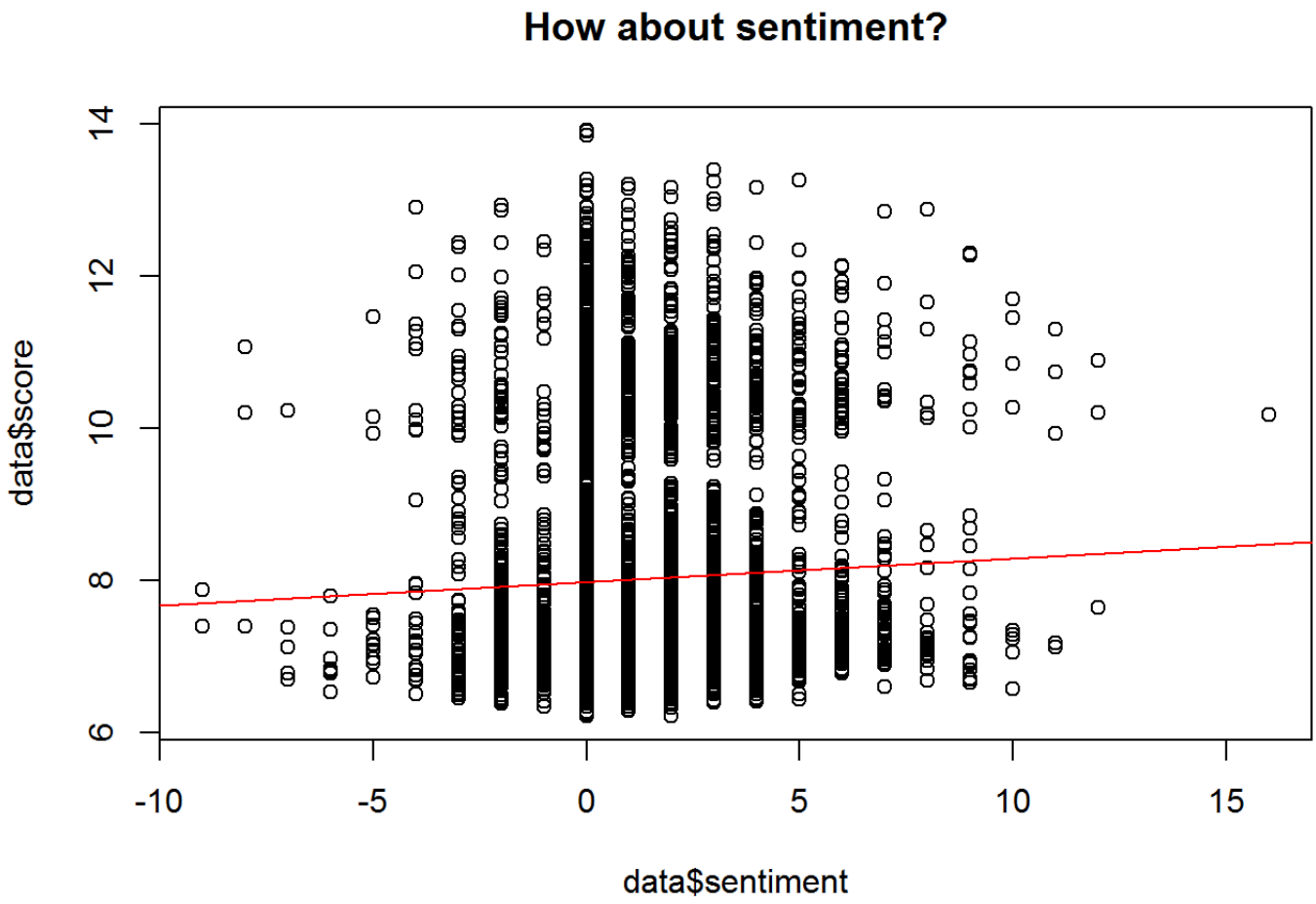
```
file = 'Tweets_Celiac_sent.csv'
data = read.csv(file)
data=data[c("listed_count","favourites_count","sentiment","retweet_count","score")]
model1 = lm(score~.,data=data)
summary(model1)
```

```
##
## Call:
## lm(formula = score ~ ., data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.392 -1.029 -0.656  0.503  5.146
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      7.92e+00   2.47e-02  320.14  < 2e-16 ***
## listed_count     4.91e-04   3.25e-05   15.13  < 2e-16 ***
## favourites_count 8.38e-06   3.58e-06    2.34  0.01937 *
## sentiment        3.26e-02   8.62e-03    3.79  0.00015 ***
## retweet_count    7.15e-03   5.00e-03    1.43  0.15254
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.51 on 5670 degrees of freedom
## Multiple R-squared:  0.0448, Adjusted R-squared:  0.0441
## F-statistic: 66.4 on 4 and 5670 DF,  p-value: <2e-16
```

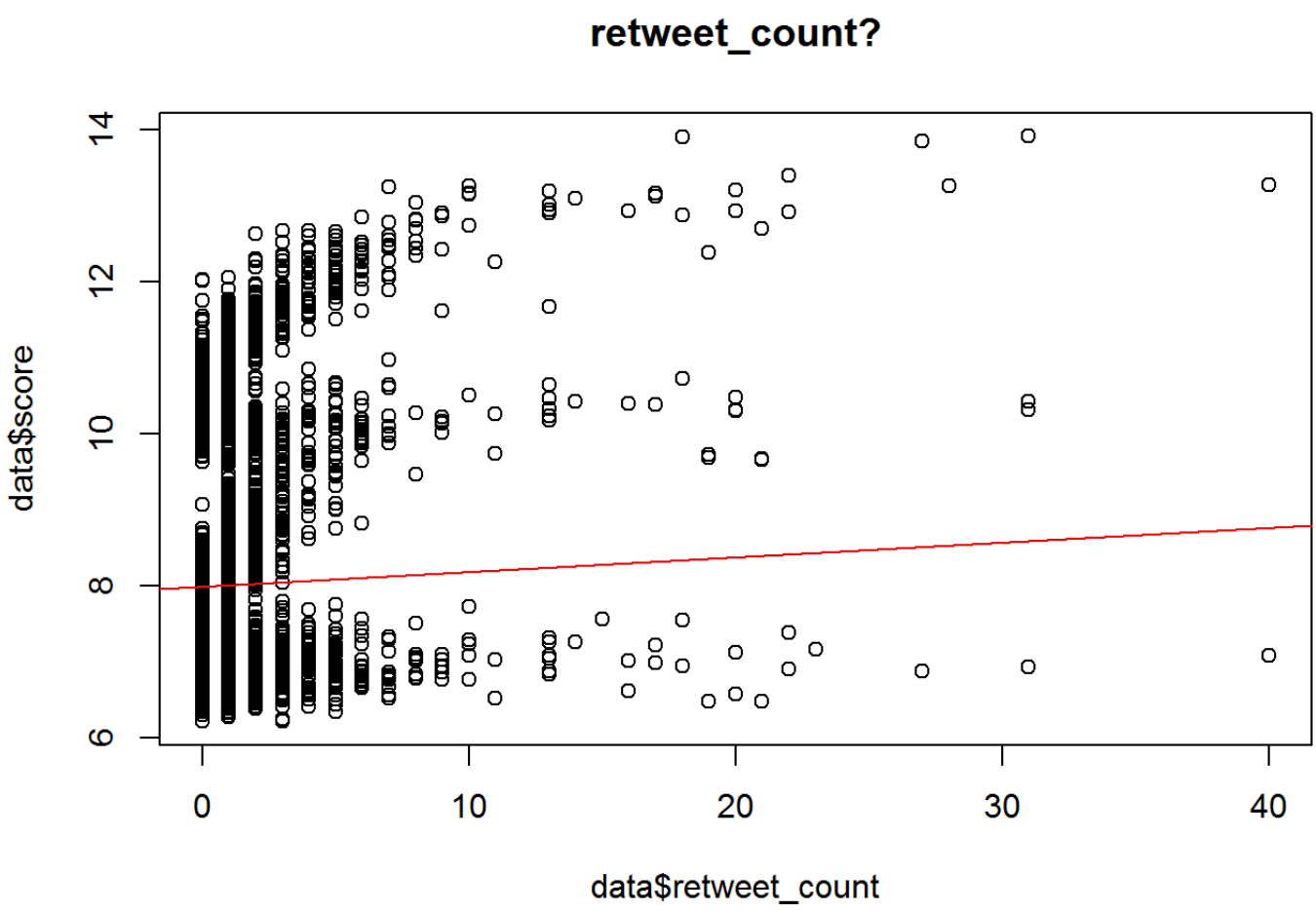
```
plot(data$listed_count,data$score,main="listed_count has the best p-value\nDoes it predict anyt
hing?")
abline(lm(score~listed_count,data=data),col='red')
```



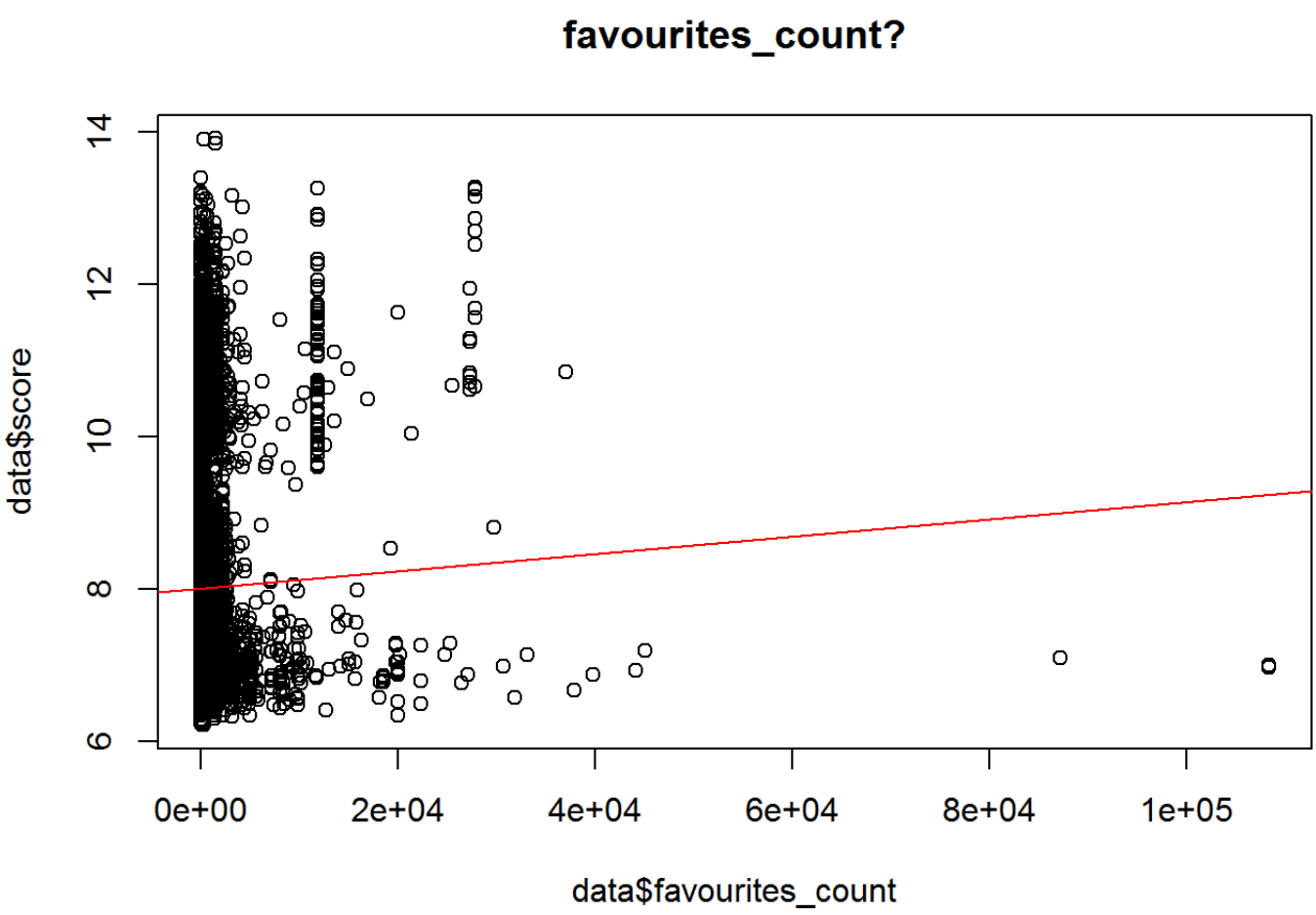

```
plot(data$sentiment,data$score,main="How about sentiment?")
abline(lm(score~sentiment,data=data),col='red')
```




```
plot(data$retweet_count,data$score,main="retweet_count?")
abline(lm(score~retweet_count,data=data),col='red')
```




```
plot(data$favourites_count,data$score,main="favourites_count?")
abline(lm(score~favourites_count,data=data),col='red')
```


favourites_count?
data$score
14
12
10
8
6
0e+00
2e+04
4e+04
6e+04
8e+04
1e+05
data$favourites_count